

# *Symmetry Arguments Against Regular Probability:*

## *A Reply to Recent Objections*

Matthew W. Parker

Centre for Philosophy of Natural and Social Science  
London School of Economics

### **Abstract**

A probability distribution is *regular* if no possible event is assigned probability zero. While some hold that probabilities should always be regular, three counter-arguments have been posed based on examples where, if regularity holds, then perfectly similar events must have different probabilities. Howson (2017) and Benci et al. (2016) have raised technical objections to these symmetry arguments, but we see here that their objections fail. Howson says that Williamson's (2007) "isomorphic" events are not in fact isomorphic, but Howson is speaking of set-theoretic representations of events in a probability model. While those sets are not isomorphic, Williamson's physical events are, in the relevant sense. Benci et al. claim that all three arguments rest on a conflation of different models, but they do not. They are founded on the premise that similar events should have the same probability in the *same* model, or in one case, on the assumption that a *single* rotation-invariant distribution is possible. Having failed to refute the symmetry arguments on such technical grounds, one could deny their implicit premises, which is a heavy cost, or adopt varying degrees of instrumentalism or pluralism about regularity, but that would not serve the project of accurately modelling chances.

## 1. Introduction

Many philosophers have suggested that probabilities (rational credences, objective chances, or both) should be regular (Carnap 1950, 1963; Kemeny 1955, 1963; Shimony 1955; Jeffreys 1961; Edwards, Lindman, and Savage 1963; De Finetti 1964; Stalnaker 1970; Lewis 1980, 1983; Skyrms 1980; Appiah 1985; Jackson 1987; Jeffrey 1992; Wenmackers and Horsten 2010; Benci et al. 2013; 2016; Hofweber 2014).<sup>1</sup> A probability measure is regular if it assigns probability zero only to impossible events, so that every strictly possible event gets positive probability. For a probability space  $\langle \Omega, F, P \rangle$ , this is represented by the condition that if  $P(A) = 0$  then  $A$  is the empty set.

Such a probability measure can be difficult to arrange, especially where the set of possible outcomes is infinite and all outcomes are equally likely, for then the regular, non-zero probabilities of these outcomes will normally add up to more than one. But we can avoid this problem if we allow  $P$  to take infinitesimal values. If we assign a non-zero infinitesimal probability to each outcome in our sample space, these need not add up to more than one.<sup>2</sup> The desideratum of regularity has been the main reason for introducing infinitesimal and hyperreal<sup>3</sup> probabilities.

---

<sup>1</sup> I do not claim that all of these authors are supporters of regularity, only that the particular works cited at least make suggestions in that direction.

<sup>2</sup> For uncountable sample spaces, regularity and finite additivity already require infinitesimals, even if the probabilities are *not* uniform. It is easy to show that, for any function  $P: \Omega \rightarrow \mathbf{R}^+$  where  $\Omega$  is uncountable and  $\mathbf{R}^+$  is the set of *positive* real numbers, there are finitely many  $\omega_0, \omega_1, \dots, \omega_n \in \Omega$  such that  $P(\omega_0) + P(\omega_1) + \dots + P(\omega_n) > 1$ .

<sup>3</sup> Hyperreal numbers are the elements of a field generated by real numbers and infinitesimals.

Some think that this desire for regularity is a naïve mistake, but the arguments for it are serious enough to warrant response. (See above references. Benci et al. 2016 reviews some of the main arguments.) Williamson (2007), Parker (2012), Benci et al. (2016), and others (Bernstein and Wattenberg 1969, Barrett 2010, Pruss 2013) have proposed arguments against regularity based on the fact that, if regularity holds, certain perfectly similar events cannot have the same probability. Howson (2017) and Benci et al. (2016) have recently tried to refute those arguments. Here we will see how their refutations go wrong.<sup>4</sup> (Thus we will buttress the case *against* a general requirement of regularity.)

We will review the three symmetry arguments that Howson and Benci et al. criticize, fleshing them out in certain respects. We will then consider and rebut Howson and Benci et al.'s objections. Both Howson and Benci et al. focus on the details of formal probability models, claiming that, once Williamson's models are made explicit, his error becomes apparent. Benci et al. claim that this extends as well to Parker's argument and their own proposal. But we will see that the technical errors alleged by Howson and Benci et al. are not present in the original symmetry arguments.<sup>5</sup> Those arguments are based on very general principles, which we will make more explicit here, and which, while they are not above doubt, are not beholden to the technical details of any particular probability model. Finally, we will consider what stances a regularist might take given that the objections fail.

---

<sup>4</sup> Easwaran (2014) defends Williamson against a different kind of objection, which we will not review here. He also argues that the expressiveness promised by regular probabilities is already provided by the non-numerical aspects of probability models, e.g., by inclusion relations between sets of outcomes.

<sup>5</sup> As we will see, Benci et al. in a sense misunderstand their own proposed argument against regularity, for they present it as a parallel to Williamson's, and in misconstruing Williamson's argument, they likewise misdiagnose their own.

## 2. The symmetry arguments

### 2.1 Williamson's coin flip argument

Consider a fair coin that is tossed infinitely many times, at times  $t_0 + n$  seconds for  $n = 0, 1, 2, \dots$ . Let  $H(1\dots) = H(1) \& H(2) \& H(3) \& \dots$  be the event that every toss comes up heads.

Williamson argues that, even if we let probabilities take hyperreal values,  $\text{Prob}(H(1\dots)) = 0$ .<sup>6</sup>

Since  $H(1\dots)$  is strictly possible, regularity fails. The crucial step in Williamson's argument is the claim that, if  $H(2\dots) = H(2) \& H(3) \& \dots$  is the event that every toss *after*  $t_0$  comes up heads, then  $H(1\dots)$  and  $H(2\dots)$  are "isomorphic events" and therefore should have the same probability.

Let  $H(1)$  be the event that the first toss comes up heads. Then

$$\text{Prob}(H(1\dots)) = \text{Prob}(H(1) \& H(2\dots)).$$

Since the coin is fair and the tosses independent, we have,

$$\text{Prob}(H(1)) = \frac{1}{2},$$

$$\text{Prob}(H(1) \& H(2\dots)) = \text{Prob}(H(1)) \times \text{Prob}(H(2\dots)),$$

and therefore,

$$\text{Prob}(H(1\dots)) = (\frac{1}{2}) \text{Prob}(H(2\dots)).$$

But since  $H(1\dots)$  and  $H(2\dots)$  are isomorphic, Williamson claims,

$$\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots)),$$

and by substitution,

$$\text{Prob}(H(1\dots)) = (\frac{1}{2}) \text{Prob}(H(1\dots)).$$

Since zero is the only solution to  $x = (\frac{1}{2})x$ ,

---

<sup>6</sup> Throughout we use 'Prob' for functions over physical events and 'P' for functions over sets that model physical events.

$$\text{Prob}(H(1\dots)) = 0.$$

Thus, the possible event  $H(1\dots)$  has probability zero, so regularity fails.<sup>7</sup>

If this argument is sound, it applies equally whether the values of Prob are real or hyperreal, since both number systems have the same first-order properties (those of a real closed field), including all the properties used in the argument.

## 2.2 Physical isomorphism

Before we review the other arguments, let us flesh out Williamson's a little. It clearly relies on the following assumption:

**Isomorphism Principle (IP):** If two events are isomorphic (in the relevant sense), they should have the same probability.

Williamson does not state IP explicitly, but what he says is suggestive:

But  $H(1\dots)$  and  $H(2\dots)$  are isomorphic events. More precisely, we can map the constituent single-toss events of  $H(1\dots)$  one-one onto the constituent single-toss events of  $H(2\dots)$  in a natural way that preserves the physical structure of the set-up just by mapping each toss to its successor.  $H(1\dots)$  and  $H(2\dots)$  are events of exactly the same qualitative type; they differ only in the inconsequential respect that  $H(2\dots)$  starts one second after  $H(1\dots)$ . Thus  $H(1\dots)$  and  $H(2\dots)$  should have the same probability.

Hence, the fact that  $H(1\dots)$  and  $H(2\dots)$  have the same qualitative physical properties is, for Williamson, the reason that they should have the same probability, and presumably this inference is undergirded by a general principle like IP.

Why should one accept IP? An argument for a version of IP might run as follows:

- (I) The laws of physics are space-time invariant.
- (II) The chance of an event is determined by the physical laws and local qualitative circumstances.

---

<sup>7</sup> Williamson introduces a third sequence of coin tosses in order to make his point more vivid, but we need not consider it here.

-----  
Therefore,

- (IP') Two events that differ at most in where and when they hypothetically occur (and perhaps in matters of bare identity but not in *qualitative* features) have the same chance.

What I mean by (I) is just that the laws of physics are the same in every place at every time, and they do not have any place- or time-dependent features. Whatever the laws imply about the outcome of an experiment is the same no matter where and when that experiment is conducted, other things being equal.<sup>8</sup> This in itself does not imply IP', because we might think that chances depend on something other than laws and qualitative circumstances. But if (II) holds as well, then IP' follows (barring any creative concept stretching).

The above argument applies directly only to physical chances or propensities. But according to the Principal Principle (Lewis 1980, 1994), our credences should generally track known chances, so we should also assign equal credence to such isomorphic events.

Thus, the regularist is in an awkward dilemma: She must either deny the standard and sensible principle that physical laws are space-time invariant, or deny that chances are determined by local circumstances and laws. Neither is inconceivable, but either is a weighty consequence, perhaps too weighty for the *a priori* arguments for regularity to sustain.

One might take the view that this argument from space-time invariance is irrelevant, since the stipulation that the individual tosses are independent and identically distributed (*iid*) already implies that  $H(1\dots)$  and  $H(2\dots)$  have the same "standard" probability.<sup>9</sup> Under the Kolmogorov axioms (including countable additivity),  $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots)) = \prod_{n \in \mathbf{N}} \frac{1}{2} = 0$  (where  $\mathbf{N} = \{0, 1, 2, \dots\}$ ). However, to appeal to that result would be begging the question, for it

---

<sup>8</sup> Notice that (I) is weaker than Galilean or Lorentzian invariance, according to which the laws of physics are the same across *all* inertial reference frames (in a Euclidean or Minkowskian space-time, respectively). Premise (I) is just the claim that, given *one* such reference frame, the laws of physics are the same and apply in the same way in each *part* of that one reference frame.

<sup>9</sup> This was argued by an anonymous reviewer.

assumes the standard axioms and number system, which regularists propose to revise. As Williamson's argument shows, the equality  $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots))$  fails under any alternative theory in which  $\text{Prob}(H(1\dots))$  and  $\text{Prob}(H(2\dots))$  are not strictly zero, since  $\text{Prob}(H(1\dots)) = \frac{1}{2} \text{Prob}(H(2\dots))$ . If one wishes to argue directly from iid to the equality of  $\text{Prob}(H(1\dots))$  and  $\text{Prob}(H(2\dots))$ , one must assume either countable additivity, which regularists reject, or some other probability axiom that regularists would likely reject, e.g., that the probability of a conjunction of independent events is entirely determined by the probabilities of the conjuncts. (This is close to what Hofweber (2014) calls conjunctive local determination and does reject.) Such a strategy fails because it assumes too much, and since it fails, the appeal to space-time invariance *is* relevant, if it succeeds. In any case, Williamson does not argue directly from iid to  $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots))$ , but appeals instead to the qualitative physical properties of the events.

### 2.3 The circle argument

Williamson's  $H(1\dots)$  and  $H(2\dots)$  are conjunctions of infinitely many coin flip outcomes, in the sense that they require infinitely many toss outcomes all to occur. (I do not wish to conflate the physical events  $H(1\dots)$  and  $H(2\dots)$  with conjunctive *sentences*.) Another argument against regularity (Bernstein and Wattenberg 1969, Barrett 2010, Parker 2012, Pruss 2013) involves disjunctive rather than conjunctive events.

Construct a set of points on the unit circle as follows: Let  $p_0 = (1, 0)$  in polar coordinates, i.e., the point on the circle due-right of the centre. Let  $p_{n+1}$  be the point  $(1, n + 1)$  on the circle, one radian counter-clockwise from  $p_n$ . Then let  $C_0 = \{p_0, p_1, p_2, \dots\} = \{(1, n) : n \in \mathbf{N}\}$ , and let  $C_1 = \{p_1, p_2, p_3, \dots\} = \{(1, n + 1) : n \in \mathbf{N}\}$ . Notice that  $C_1$  is a rotation of  $C_0$  by one radian, but is also a proper subset of  $C_0$ , since  $C_1$  does not contain  $p_0$ .

Now, let us choose a point on the circle randomly, say by throwing a dart at the interior disk and constructing a radius through the centre of the dart shaft to a point on the circle. What is the probability that this point lies in  $C_0$ , and what is the probability that it lies in  $C_1$ ? We can

model this experiment with a probability space  $\langle S^1, F, P \rangle$ , where  $S^1$  is the unit circle and  $F$  an algebra on subsets of the circle that at least include  $C_0, C_1$ , and the singleton  $\{p_0\}$ . The event  $E_C$  that the point chosen by our experiment lies in a given set  $C$  is modelled by that very set, i.e.,  $\text{Prob}(E_C) = P(C)$ , where  $\text{Prob}$  is the chance or credence of a physical occurrence and  $P$  is a function on sets that models  $\text{Prob}$ . Thus, in our model, the set  $C_0$  represents the disjunctive event that the point chosen by the dart throw is  $p_0$  or  $p_1$  or  $p_2$  or ... .

Now, *assume*<sup>10</sup> that  $P$  is rotationally symmetric. Then  $P(C_0) = P(C_1)$ . But by finite additivity,  $P(C_0) = P(\{p_0\}) + P(C_1)$ . Hence,  $P(\{p_0\}) = 0$ , contradicting regularity. And as with Williamson's argument, this holds whether  $P$  takes hyperreal values or only real values.

There are significant differences between this argument and Williamson's. Firstly, the circle experiment takes place in a finite region of space-time. It is just a single dart throw at a finitely bounded disc (or in other versions, a single spin of a spinner or a single quantum vacuum fluctuation). Thus it avoids Williamson's unrealistic hypothesis of an eternal sequence of tosses, in perfect rhythm, of a single, ever unchanging coin. And if one is tempted to dodge Williamson's argument by suggesting that space-time invariance only applies to finite experiments and not to temporally infinite sequences of events, such a dodge will not escape the circle argument.

Secondly, the circle argument does not rely on IP.<sup>11</sup> It simply assumes that the distribution is rotationally symmetric. However, this is only plausible if a dart throw or some

---

<sup>10</sup> I do not claim that  $P$  *must* be rotationally symmetric. That would perhaps be question begging, but in any case mistaken. Real dart throws are presumably not perfectly rotationally symmetric. I merely claim that it is *plausible* that a rotationally symmetric distribution is *possible*. Committed regularists would have us believe that a possible event *must always* have positive probability. This argument shows that such a claim has a rather strong consequence, namely that fully rotationally symmetric continuous distributions are *impossible*.

<sup>11</sup> Thus, not all of the symmetry arguments depend on space-time invariance, nor on the assumption that chances are determined by laws and local circumstances.



other experiment really can be performed with a perfectly symmetric distribution.<sup>12</sup> Intuitively this ought to be possible, but to my knowledge there is no standard physical principle to guarantee it in the way that space-time invariance (along with (II)) guarantees Williamson's result. It does not help much to appeal to empirically confirmed laws that imply symmetry, for a committed regularist will claim that our empirical laws need a slight revision, so that any probabilities that they imply are adjusted by infinitesimal amounts to make them regular. Such subtle revisions are generally compatible with observed frequencies. If we could construct an example where the symmetry is due to some general principle rather than specific laws, the regularists would not have such an easy retort. Parker 2012 attempts to construct such an example involving quantum vacuum fluctuations, but the success of that example is debatable.<sup>13</sup> So an uncontroversially realistic and principled example is yet to be given but is far from being ruled out.<sup>14</sup> Furthermore, Benci et al. hold that a probability theory ought to be able to describe

---

<sup>12</sup> In the circle argument stated here, we require rotational symmetry, but as shown in Parker 2012, there are transformations of the circle example where translation or reflection symmetry is sufficient.

<sup>13</sup> The example relies on the familiar picture of quantum vacuum fluctuations, in which virtual particles appear and annihilate each other independently of observations. Some have argued in another context that this picture is naïve (Boddy, Carroll, and Pollack 2016, 2017).

<sup>14</sup> An anonymous reviewer suggests that none of our examples is physically realistic in any world whose laws are even remotely like our world's, and questions whether we can trust our intuitions about such examples. I am not certain that the examples are so unrealistic. The dart throw and the related vacuum fluctuation of Parker 2012 only require that exact real values are selected from a continuum in a rotation- or translation-invariant way. Quantum mechanics and indeed common sense suggest that we cannot measure such exact values, but they do not imply that no exact values exist in the world. If a quantum vacuum fluctuation does not in fact define any exact point in space and time, perhaps the exact center of some well defined lump in the Schrödinger wave function does. We need not be able to measure an exact value in order to argue that it is, or could be, determined by an invariant distribution and therefore contradicts regularity. Furthermore, if we must imagine another world in which such a distribution is

conceptually possible processes such as a fair lottery with infinitely many tickets, and a dart throw with a perfectly symmetric distribution seems at least as conceivable as a fair infinite lottery.<sup>15</sup> So if we need a probability theory that makes sense of the infinite lottery, as Benci et al. claim, then we arguably need one that also accommodates invariant continuous distributions. But the circle argument shows that such a theory cannot be regular.

## 2.4 The urn argument

Benci et al. (2016) introduce their own symmetry argument against regularity, intending to refute it and thereby illustrate how the other ones go wrong. Their argument is based on a fair infinite lottery, which is also the main motivating example for their Non-Archimedean Probability (NAP) theory (2011, 2016; cf. Wenmackers 2011, Wenmackers and Horsten 2013). Their argument runs as follows:

Imagine an urn containing a countably infinite collection of tickets and a mechanism to implement a fair lottery on the tickets in the urn.

In situation (1), all tickets are in the urn and we denote the probability of winning of each arbitrary single ticket in such a lottery as  $\text{Prob}(E_1)$ , leaving open the possibility that this may be an infinitesimal.

In situation (2), one ticket is removed from the urn prior to the drawing of the winning ticket. There is one competing ticket less, so the probability of winning of each remaining ticket is  $\text{Prob}(E_2) = \frac{1}{1 - \text{Prob}(E_1)} \text{Prob}(E_1)$  (renormalization). Taken in isolation, however, situation (2) looks exactly as before the removal of a ticket, which is situation (1). Because of this isomorphism between situation (1) and situation (2), we find that the probability of winning of each individual ticket is equal to  $\text{Prob}(E_2) = \text{Prob}(E_1)$ . ... Even in a non-Archimedean [hyperreal] field, these equalities can only hold simultaneously if  $\text{Prob}(E_1) = \text{Prob}(E_2) = 0$ . (Benci et al. 2016)

Thus, according to this argument,  $E_1$  and  $E_2$  are possible but have probability zero, so again regularity fails.

---

possible, we do not need any further intuition about the physics of the process in order to conclude that it would violate regularity.

<sup>15</sup> It is not obvious that a fair infinite lottery really is conceptually coherent, but Benci et al. take it to be so.

The relevant “isomorphism” here is expressed in the stipulation that the new situation (2) “looks exactly as before”. The qualitative physical circumstances are the same, or at least, the argument assumes that they are sufficiently alike that the probability of choosing a given ticket should be the same in situations (1) and (2). If we like, we can further stipulate that the remaining tickets in situation (2) shift so that they have exactly the same states as those in situation (1). Then, as in the coin argument, IP implies that the probabilities are the same.

## 2.5 A problem with the urn argument

Below we will turn to Benci et al.’s attempt to refute this argument, but let us note here a problem with it that they do not discuss. The renormalization step, according to which  $\text{Prob}(E_2) = \frac{1}{1 - \text{Prob}(E_1)} \text{Prob}(E_1)$ , is not obviously correct. It assumes that removing a ticket increases the probability of being selected for each remaining ticket, but this need not be so. Removing a ticket changes the physical situation, at least in terms of bare identities, and for the regularist there is no general rule that says this will change the probability for any of the remaining tickets. Regularity does imply that removing a ticket increases the probability that the chosen ticket will lie in the *set* of all other tickets,<sup>16</sup> but that does not imply that the probability has changed for any *particular* ticket. It may be tempting to object, “How can the probability increase for the set of remaining tickets if it does not increase for any individual ticket?”, but that objection presupposes countable additivity, or something like it, and proponents of infinitesimal probabilities are already willing to sacrifice countable additivity (e.g., Benci et al. 2013; 2016).

One might think that the renormalization step is justified by conditionalization, as Benci et al. later seem to suggest (p. 19). Let  $C_t$  be the event that a ticket  $t$  is chosen in situation (1).

---

<sup>16</sup> Let  $T$  be the set of tickets and  $t$  the ticket removed. In situation (1), the probability that a ticket in  $T \setminus \{t\}$  is selected is  $1 - \text{Prob}(E_1)$ . Assuming regularity and that  $\text{range}(\text{Prob})$  is contained in an ordered field,  $1 - \text{Prob}(E_1) < 1$ . In situation (2), the probability that a ticket in  $T \setminus \{t\}$  is selected is one, which is larger.

Since the lottery is fair,  $\text{Prob}(C_t) = \text{Prob}(C_u) = \text{Prob}(E_1)$  for any two tickets  $t$  and  $u$ . Now suppose that ticket  $t$  is the one removed in situation (2). If we therefore assume that  $\text{Prob}(E_2)$  is equal to the conditional probability  $\text{Prob}(C_u \mid \sim C_t)$ , the ratio formula for conditional probability gives us

$$\text{Prob}(E_2) = \frac{\text{Prob}(C_u \& \sim C_t)}{\text{Prob}(\sim C_t)} = \frac{\text{Prob}(C_u)}{1 - \text{Prob}(C_t)} = \frac{1}{1 - \text{Prob}(E_1)} \text{Prob}(E_1),$$

which is precisely the renormalization step. But this assumption that  $\text{Prob}(E_2) = \text{Prob}(C_u \mid \sim C_t)$  is not obviously correct either. Conditional probability is commonly used to model situations where we have obtained some information about an outcome, such as the news that ticket  $t$  was not chosen. It also models cases where we adopt a policy, e.g., if ticket  $t$  is chosen, put it back and repeat the experiment. But Benci et al.'s case is neither of those. It is a case where a ticket has been physically removed, and there is no general rule about how that will affect the probabilities for the remaining tickets. So this application of conditional probability is unjustified.

Thus, the argument that regularity contradicts IP in this case is incomplete. However, there could conceivably be situations, perhaps specific selection mechanisms, for which the renormalization step or some similar move<sup>17</sup> is correct. As with the circle argument, such a situation is at least as conceivable as the fair infinite lottery itself. But regardless of whether the urn argument can be salvaged, we will see that Benci et al.'s way of countering it, as with the other arguments, is unsuccessful.

### 3. Objections and replies

#### 3.1 Howson's objection to Williamson

---

<sup>17</sup> Benci et al.'s renormalization step implies that the probabilities for all remaining tickets are affected equally, i.e., multiplied by the same factor, but that is not needed. All that is needed to complete the urn argument is a case where removing a ticket multiplies the probability for *some* particular ticket by a factor other than one.

Howson claims that Williamson's argument fails due to "a confusion about what he calls 'isomorphic events', assisted by an inadequate notation." The key point in Howson's argument is that, in an appropriate probability model for Williamson's example,  $H(1\dots)$  is a singleton set, containing just one element of the sample space, while  $H(2\dots)$  is a pair, containing two elements of the sample space. Since a singleton is not isomorphic to a pair, the events are not isomorphic and the argument fails. Howson goes on to consider variations on Williamson's argument, but this is the main thrust of his objection.

To be precise, Howson specifies a probability space  $\langle 2^{\mathbb{N}}, F, P \rangle$ , where the sample space  $2^{\mathbb{N}}$  is the set of all countable, one-way infinite sequences of zeros and ones, such as  $\langle 0, 1, 1, 0, \dots \rangle$ ;  $F$  is the algebra generated by the cylinder sets<sup>18</sup> of  $2^{\mathbb{N}}$ ; and  $P$  is a hyperreal-valued probability function defined on  $F$ . Williamson's event  $H(1\dots)$  is then modelled as the set  $\{\langle 1, 1, 1, \dots \rangle\}$  and  $H(2\dots)$  as the set  $\{\langle 1, 1, 1, \dots \rangle, \langle 0, 1, 1, \dots \rangle\}$ . Note that the elements of an "event" in probability theory represent disjunctive alternatives. The sets  $\{\langle 1, 1, 1, \dots \rangle, \langle 0, 1, 1, \dots \rangle\}$ , for example, correspond to the disjunction  $([H(1) \& H(2) \& H(3) \& \dots] \text{ or } [T(1) \& H(2) \& H(3) \& \dots])$ , where  $T(1)$  is the event that the first outcome is tails.

In general, two collections are said to be isomorphic if there is a bijection between them that preserves all relevant structure. In algebra, for example, an isomorphism preserves the algebraic relations between elements. But the sets  $\{\langle 1, 1, 1, \dots \rangle\}$  and  $\{\langle 1, 1, 1, \dots \rangle, \langle 0, 1, 1, \dots \rangle\}$  are not isomorphic in any sense, because there is not even a bijection between them. Thus, according to Howson, Williamson's events are not isomorphic, so his argument is a non-starter.

### 3.2 Reply

---

<sup>18</sup> The assumption that  $F$  is an algebra generated by cylinder sets is unnecessary here. If we are willing to relinquish the possibility of translation invariance, as regularists must, we can define  $P$  on the entire power set of  $2^{\mathbb{N}}$ . But the particular domain of  $P$  does not bear on Howson's objection so long as it includes  $\{\langle 1, 1, 1, \dots \rangle\}$  and  $\{\langle 1, 1, 1, \dots \rangle, \langle 0, 1, 1, \dots \rangle\}$ .

What Williamson means by “isomorphic events” does not concern “events” in the jargon of probability theory, i.e., sets in the algebra of a probability space, but physical *events*, in the ordinary sense of *things that happen*, or things that *might* happen. As we saw, Williamson is concerned with “the physical structure of the set-up” and the “qualitative type” of events. Moreover, he explains what he means by ‘isomorphic events’ in terms of a structure-preserving map, not between subsets of the sample space, but between “the constituent single-toss events”  $H(1), H(2), H(3), \dots$  and  $H(2), H(3), H(4), \dots$  that make up the events  $H(1\dots)$  and  $H(2\dots)$ , respectively. While Howson’s “events” are effectively sets of disjuncts, Williamson’s “isomorphism” consists in a mapping between the *conjunct* events in  $H(1\dots) = H(1) \& H(2) \& H(3) \& \dots$  and  $H(2\dots) = H(2) \& H(3) \& H(4) \& \dots$  (I do not mean a mapping between the symbols that serve as conjuncts in a sentence, but between the physical events that form a larger event.) Furthermore, Williamson’s set-up guarantees a mapping of the conjunct events that preserves qualitative physical properties and relations. He specifically stipulates that the coin tosses in his sequences use the same fair coin and that the time intervals between tosses are the same. We could go further and stipulate that *all* qualitative properties of the tosses and sequences of tosses are exactly the same. Then the events  $H(1\dots)$  and  $H(2\dots)$ , construed not as subsets of a sample space but as physical things that might happen, are indeed isomorphic in Williamson’s intended sense. On his view,  $P(\{ \langle 1, 1, 1, \dots \rangle \})$  should equal  $P(\{ \langle 1, 1, 1, \dots \rangle, \langle 0, 1, 1, \dots \rangle \})$ , not because these two *sets* are isomorphic (which they clearly are not), but because the physical events they model are qualitatively alike.<sup>19</sup>

---

<sup>19</sup> An anonymous reviewer objects that  $H(1\dots)$  and  $H(2\dots)$  are *not* qualitatively identical because  $H(2\dots)$  accommodates the possibility of a tail at the first flip and  $H(1\dots)$  does not. But as Williamson understands these events, neither accommodates a tail at “the first flip”. Let us label the flips of  $H(1\dots)$  flip 1, flip 2, etc. Then on Williamson’s conception, the “first flip” of  $H(2\dots)$  is flip 2, which must come up heads if  $H(2\dots)$  occurs. The reviewer’s objection seems to require that  $H(2\dots)$  somehow incorporate the outcome of flip 1, even though flip 1 is logically and probabilistically independent of the occurrence or non-occurrence of  $H(2\dots)$ . I think that Williamson, on the other hand, thinks of  $H(2\dots)$  as consisting in certain *intrinsic*

Furthermore, there is an argument available that such isomorphism ought to imply equal probability, namely the argument above from (I) and (II), and this argument appeals only to the physical character of the events-in-the-ordinary-sense, not to the set-theoretic structure of the sets that represent those events in a particular mathematical model. Whatever model we might adopt, the physical argument still holds sway, so far as its premises are plausible. So Howson's set-theoretic objection misses the mark by a wide margin.

To be fair, Williamson's use of 'isomorphism' to express physical similarity shoulders some blame for this misunderstanding. An isomorphism is normally a mapping between sets, not between physical things-that-could-happen. Nonetheless, the intended principle is clear: If the same experiment is conducted under the same conditions at two different times, the probabilities of the outcomes should be the same. Howson seems to have ignored this point in order to raise a technical issue that, in the end, is not relevant.

### 3.3 Events as sentences

Just as Williamson's "isomorphic events" are not sets, they are not sentences either.<sup>20</sup> They are physical things-that-might-happen. Thus, as noted above, when we speak of a mapping between the conjuncts that constitute  $H(1\dots)$  and  $H(2\dots)$ , this is not a mapping between the symbolic conjuncts in a sentence, but between the physical coin flip events that make up  $H(1\dots)$  and  $H(2\dots)$ . Otherwise, one might be tempted to object that  $H(1\dots)$  and  $H(2\dots)$

---

structural properties of the physical motion of a coin beginning at time  $t_0 + 1$ , and not concerning in any way events occurring before  $t_0 + 1$ . Thus,  $H(2\dots)$  does not at all concern flip 1, which occurs at time  $t_0$ . Likewise  $H(1\dots)$ , for Williamson, concerns the intrinsic structural properties of the motion of a coin from time  $t_0$  onward, and nothing else. Williamson's isomorphism claim is just the claim that these structural properties, those of the coin's motion from  $t_0$  onwards in  $H(1\dots)$  and those of its motion from  $t_0 + 1$  onwards in  $H(2\dots)$ , are qualitatively identical, and this is just true by stipulation.

<sup>20</sup> This interpretation was suggested by an anonymous reviewer.

cannot be isomorphic because the first is just the infinite *sentence* 'H(1) & H(2) & H(3) & ...' while the second is an infinite sentence of the disjunctive form

$H(2\ldots)_{\text{Dis}}$ : 'H(1) & H(2) & H(3) & ... *or* T(1) & H(2) & H(3) & ...'.

To do so would miss two important points.

First, the fact that  $H(2\ldots)$  can be viewed as a disjunction or disjunctive event does not imply that it must be so understood. To say that  $H(2\ldots)$  is *really* of the form  $H(2\ldots)_{\text{Dis}}$  is like saying that the event of my winning the lottery today is *really* the event of my winning the lottery today and having tuna for lunch *or* winning the lottery today and *not* having tuna for lunch. The tuna is a red herring. It is irrelevant to my winning the lottery, just as the first coin flip in  $H(1\ldots)$  is irrelevant to the probability of  $H(2\ldots)$ . Even if we were to understand  $H(1\ldots)$  and  $H(2\ldots)$  as sentences, there is no reason that  $H(2\ldots)$  must take the form  $H(2\ldots)_{\text{Dis}}$ , for, given Williamson's specification of the experiments,  $H(2\ldots)_{\text{Dis}}$  is logically equivalent to the simpler sentence 'H(2) & H(3) & H(3) ...', which is structurally identical to the natural expression 'H(1) & H(2) & H(3) & ...' for  $H(1\ldots)$ .

But this brings us to the second point: Williamson is not concerned with a mapping between linguistic objects, but between possible physical occurrences. It is not a structural similarity between symbol strings that concerns him, but between physical events that might occur in the world. If you like, these can be understood as mathematical structures that might be instantiated by the behaviour of physical coins. They are not instantiated by letters on paper or words from a mouth. Again, this is clear from Williamson's deliberate specification of the physical conditions and his references to "physical structure" and "qualitative type". The important thing for Williamson is that, by stipulation, there is no intrinsic qualitative difference between the physical events  $H(1\ldots)$  and  $H(2\ldots)$ . They differ in their times of occurrence, their relations to each other and other events, and their bare haecceities if you like, but those are precisely the kinds of things that, on the plausible principle IP, make no difference to their probabilities.



### 3.4 Howson, the circle, and the urn

Howson 2017 is concerned only with the coin flip argument and a couple of variations on it, not with the circle argument or the urn. However, it is worth noting that, while Howson's critique of the coin flip argument misses the point, it does not apply at all to the other arguments. In the circle argument, the physical event that the point chosen by the dart throw lies in  $C_0$  is represented by the set  $C_0$  itself, and likewise for  $C_1$ . Since  $C_1$  is just a rotation of  $C_0$ , the two sets are set-theoretically, topologically, and metrically isomorphic. But the circle argument does not rely on isomorphism per se. It uses the fact that the two sets are rotations of each other, and assumes explicitly that the probability distribution is rotationally symmetric. So Howson's complaint about Williamson's argument, that the sets representing the events are not isomorphic, is not true of the circle argument, nor relevant.

Nor is there a parallel of Howson's critique for the urn argument. There the event of drawing a given ticket is naturally represented by a singleton  $\{n\} \subseteq \mathbf{N}$ , in both situations (1) and (2). Since any such singletons are isomorphic, the set-theoretic "events" in the models are indeed isomorphic and Howson's objection does not apply. Of course, the fact that two singletons are trivially isomorphic is no reason that they should be assigned the same probability, but the urn argument does not rely on such an isomorphism between subsets of the sample space. Like the coin flip argument, it turns instead on a qualitative similarity between physical situations. So here too, an objection along Howson's line is both false and irrelevant.

### 3.5 Benci et al.'s objection to the urn argument

Benci et al. point out that we can model the urn experiments in different ways, and while we can equate the probability of  $E_1$  under one model with that of  $E_2$  under another, we should not normally compare the probabilities given by two different models or interpretations. "[C]hanging the sample space mid-game," as they put it, "is, in general, not allowed."

Specifically, they suggest that we model situation (1) with the sample space  $\mathbf{N}$  and a hyperreal-valued function  $P$  on the subsets of  $\mathbf{N}$  such that  $P(\{n\}) = 1/\alpha$  for each  $n \in \mathbf{N}$ , where  $\alpha$  is the size of  $\mathbf{N}$  in a non-standard measure (“numerosity”).<sup>21</sup> Benci et al. call this model  $A$  and write  $\text{Prob}_A(E_1)$  for the probability that a given ticket is selected in situation (1) under model  $A$ .<sup>22</sup> According to them, we can represent situation (2) in the same model using conditional probability: The probability of selecting ticket number  $n$  given that some other ticket  $i$  has been removed is given, they claim, by

$$\text{Prob}_A(E_2) = P(\{n\}|\mathbf{N}\setminus\{i\}) = 1/(\alpha - 1) > 1/\alpha = \text{Prob}_A(E_1).^{23}$$

So, on model  $A$ , the probability that a given ticket is chosen varies from situation (1) to (2).

On the other hand, they point out, we can also represent situation (2) by letting  $\mathbf{N}$  represent the *remaining* tickets, after ticket  $i$  has been removed. In that case, a natural model  $B$  gives the probability  $\text{Prob}_B(E_2) = 1/\alpha$  of choosing one of the remaining tickets in situation (2). Thus,  $\text{Prob}_A(E_1)$  is equal to  $\text{Prob}_B(E_2)$ , but, according to Benci et al.,  $A$  and  $B$  are two different models, and the fact that two events have the same probability under two different models does not imply that they must have the same probability in a single model.

---

<sup>21</sup> Benci and others (e.g., Benci 1995; Benci and Di Nasso 2003; Di Nasso and Forti 2010; Benci, Bottazzi and Di Nasso 2014) have developed an alternative theory of set size called numerosity theory, where the size of an infinite set is not a Cantorian cardinal number but a hyper-integer. From this, one can derive a fair NAP distribution by assigning to each element of a sample space the hyperreal probability equal to the reciprocal of the numerosity of the sample space.

<sup>22</sup> Remember,  $\text{Prob}_A$  is not identical to  $P$ .  $\text{Prob}_A$  applies to physical events, while  $P$  applies to subsets of  $\mathbf{N}$ . If we let  $l_A$  be a “labelling” that maps physical events to sets in  $\mathbf{N}$ , then  $\text{Prob}_A$  can be understood as the composition  $P \circ l_A$ , i.e.,  $\text{Prob}_A(E) = P(l_A(E))$ .

<sup>23</sup> As noted above, this is not obviously correct, since in situation (2) we are not merely conditionalizing but considering an altered setup, and we cannot assume countable additivity here.

Let us clarify something here: Technically,  $A$  and  $B$  are not different models. They are two names for the same mathematical model, with the same sample space  $\mathbf{N}$  and the same assignments of values to subsets of  $\mathbf{N}$  (or at least, Benci et al. do not indicate any difference between the assignments). However, this model is associated with two different interpretations or “labellings”, different ways of associating physical events with sets in the model. For “model  $A$ ”, each natural number corresponds to one of the original tickets, while for “model  $B$ ”, each natural number corresponds to one of the *remaining* tickets, after one has been removed. So the real difference is not between two models but between two labellings. However, Benci et al.’s point is no less valid; the fact that two events have the same probability under different labellings does not imply that they must have the same probability under a single labelling.

Thus, according to Benci et al., the urn argument commits an oversight. We thought we had shown simply that  $\text{Prob}(E_1) = \text{Prob}(E_2)$ , when actually we had only shown that  $\text{Prob}_A(E_1) = \text{Prob}_B(E_2)$ , for two different labellings  $A$  and  $B$ , and this does not support the conclusion that regularity fails.

### 3.6 Reply

Benci et al. are quite right: The fact that two events have the same probability under two different labellings does not imply that they simply have the same probability. However, there is a further reason to think that, under any *accurate* model, the probability of drawing a given ticket in situations (1) and (2) should be the same. The reason is that the qualitative physical situation is exactly the same in both cases, and by our principle IP, the same event under the same qualitative circumstances should have the same probability. Benci et al. even make such an argument themselves when they write that “situation (2) *looks exactly as before* the removal of a ticket,... *Because of this isomorphism* between situation (1) and situation (2), we find that the probability of winning of each individual ticket is equal” (2016, my emphasis). Yet, when they come to their reply, they ignore the premise that “isomorphic” events should have the same probability and claim instead that the argument trades on a conflation of two different

labellings. In fact, their own presentation of the argument involves no such conflation; it is clearly founded on IP, and as noted, there is an argument for IP from more basic principles. Given IP, any model in which the probability of drawing a given ticket (other than the one removed) varies from situation (1) to situation (2) is an *inaccurate* model. Benci et al. have tried to show that one can construct such a model, where  $\text{Prob}_A(E_2) > \text{Prob}_A(E_1)$ , but that does nothing to refute the argument from IP that any such model is inaccurate.

### 3.7 Benci et al.'s objection to the coin flip argument

Following the same line as their objection to the urn argument, Benci et al. claim that Williamson conflates two different probability models for his coin flip experiments. They again refer to “two models”  $A$  and  $B$ , which are actually the same model with two different labellings. In model  $A$ , they say, the sample space (or more accurately the labelling) “reflects that the count of events starts at the first toss of  $H(1\dots)$ ”, while in model  $B$  the same sample space (with a different labelling) is used “to reflect that the count of events starts at the first toss of  $H(2\dots)$ .”

Let us make this more explicit. Define labellings  $l_A$  and  $l_B$  so that

$$\begin{aligned} l_A(H(1\dots)) &= \{\langle 1, 1, 1, \dots \rangle\}, \\ l_A(H(2\dots)) &= \{\langle 0, 1, 1, \dots \rangle, \langle 1, 1, 1, \dots \rangle\}, \\ l_B(H(2\dots)) &= \{\langle 1, 1, 1, \dots \rangle\}. \end{aligned}$$

Thus, under  $l_A$ ,  $H(1\dots)$  and  $H(2\dots)$  are represented by the same sets as in Howson's objection, while under  $l_B$ ,  $H(2\dots)$  is represented by  $\{\langle 1, 1, 1, \dots \rangle\}$  and  $H(1\dots)$  has no representation at all. Now let  $P: 2^{\mathbb{N}} \rightarrow [0, 1]^*$  where  $[0, 1]^*$  is a hyperreal unit interval, and for any physical event in the domain of  $l_X$ , for  $X = A, B$ , let  $\text{Prob}_X(E) = P(l_X(E))$ . Hence,

$$\begin{aligned} \text{Prob}_A(H(1\dots)) &= P(l_A(H(1\dots))) = P(\{\langle 1, 1, 1, \dots \rangle\}), \\ \text{Prob}_A(H(2\dots)) &= P(\{\langle 0, 1, 1, \dots \rangle, \langle 1, 1, 1, \dots \rangle\}), \text{ and} \\ \text{Prob}_B(H(2\dots)) &= P(\{\langle 1, 1, 1, \dots \rangle\}) = \text{Prob}_A(H(1\dots)). \end{aligned}$$

Now Benci et al. write,

Williamson exploits the intuition that  $\text{Prob}_A(H(1\dots)) = \text{Prob}_B(H(2\dots))$ . But he glosses this as  $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots))$ , thus turning the probabilities involved into evaluations within the same model. On the other hand, Williamson convincingly argues that  $\text{Prob}(H(1\dots)) = \frac{1}{2} \text{Prob}(H(2\dots))$ . ... The two glosses indeed contradict each other unless  $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots)) = 0$ . But the contradiction can only be obtained when the difference between the sample spaces is glossed over. (2016, 22)

In other words, Williamson's claim that  $H(1\dots)$  and  $H(2\dots)$  should have the same probability is founded on a conflation of  $\text{Prob}_A$  and  $\text{Prob}_B$ .

### 3.8 Reply

There is no textual evidence that Williamson conflates two different models or labellings. In fact, there is clear evidence to the contrary: Williamson *tells* us why  $H(1\dots)$  and  $H(2\dots)$  should have the same probability. It is because they are (physically) isomorphic. That argument does not depend on the particular model or labelling employed. Williamson's point is that, in any *accurate* model of his proposed experiment,  $H(1\dots)$  and  $H(2\dots)$  will have the same probability. Thus he would insist, not that  $\text{Prob}_A(H(1\dots)) = \text{Prob}_B(H(2\dots))$ , but that  $\text{Prob}_A(H(1\dots))$  *should* equal  $\text{Prob}_A(H(2\dots))$ , or else  $A$  is just not a good model. There is no reason to suppose that this is founded on any slide or conflation. It is clearly founded on the principle that physically isomorphic events should have the same probability, and again, there is a simple argument available for that principle. Thus, contrary to Benci et al.'s claim in the above passage, the contradiction can indeed be obtained in a single model, with a single sample space, if one only takes seriously Williamson's premise that physically isomorphic events have the same probability.

Later (p. 38), Benci et al. acknowledge the physical basis of Williamson's argument, writing, "We know, one might say, that the laws of physics are time-translation invariant." Yet they then complain that "it is still not easy to see why the NAP treatment of Williamson's scenario has to violate time-translation invariance." Well, the reasons are straightforward:

1. Any *regular* probability model that assigns probabilities to  $H(1\dots)$  and  $H(2\dots)$  must assign a larger probability to  $H(2\dots)$ .
  2.  $H(2\dots)$  is a time translation of  $H(1\dots)$ .<sup>24</sup>
  3. Therefore, any regular probability model for  $H(1\dots)$  and  $H(2\dots)$  violates time translation invariance.
  4. All NAP models are regular.
  5. Therefore, any NAP model for  $H(1\dots)$  and  $H(2\dots)$  violates time translation invariance.
- QED.

It is also true that, given a NAP model that assigns a probability to  $H(1\dots)$ , one could consider a different NAP model that assigns the same probability to  $H(2\dots)$ , but that is irrelevant. The first model on its own must violate time invariance because it assigns a larger probability to  $H(2\dots)$  than  $H(1\dots)$ , and the second model must *also* violate time invariance because it assigns a larger probability to  $H(3\dots)$  (the event that each flip after the second comes up heads) than to  $H(2\dots)$ . Moreover, if we want to understand the relation between the probabilities of  $H(1\dots)$  and  $H(2\dots)$ , we need to represent them together in one model. If that model is regular, it cannot be time invariant.

### 3.9 Benci et al.'s objection to the circle argument

Benci et al. rehearse a version of the circle example, referring to Parker 2013 and others.<sup>25</sup> They then remark, "It will be clear to the reader by now that our diagnosis of the argument from

---

<sup>24</sup> An anonymous reviewer objects that  $H(2\dots)$  is not a time translation of  $H(1\dots)$  because the former accommodates the possibility of a tail at the first flip and the latter does not. But as argued in note 18,  $H(2\dots)$  can be understood, as Williamson seems to understand it, as a physical event that occurs entirely after the first flip of  $H(1\dots)$  and does not involve that first flip *at all*.

<sup>25</sup> The example in Parker 2013 is suggestive but does not concern probability. Rather it is used to argue that "Euclidean" theories of cardinality such as numerosity (see note 20) also violate rotation and

rotational symmetry against infinitesimal probabilities is structurally identical to our diagnosis of Williamson's argument. Hence, we do not describe it in detail here."

So let us describe it in detail. The diagnosis of Williamson's argument was that he conflates two different probability models, or more precisely, two different labellings. Presumably, then, Benci et al. would claim that the circle argument tacitly appeals to two different labellings  $I_A$  and  $I_B$ , such that  $\text{Prob}_A(C_0) = P(I_A(C_0)) = P(I_B(C_1)) = \text{Prob}_B(C_1)$ . Then they will say (if the diagnosis is indeed structurally identical to that of Williamson's argument) that the circle argument tacitly switches labellings mid-game, and if we do not conflate  $\text{Prob}_A$  with  $\text{Prob}_B$ , there is no reason to suppose that  $\text{Prob}_X(C_0) = \text{Prob}_X(C_1)$  on any one labelling  $I_X$ .

### 3.10 Reply

In reality, the circle argument stated here explicitly assumes that there is a single finitely additive (and possibly hyperreal) probability function  $P$  that assigns values to both  $C_0$  and  $C_1$  and which is rotationally symmetric.<sup>26</sup> It follows trivially that  $P(C_0) = P(C_1)$ , because  $C_1$  is a rotation of  $C_0$ . And, as noted, the event  $E_C$  that the point determined by a dart throw lies in a set  $C$  is represented by that very set  $C$ . So there is only one labelling in play, namely  $I(E_C) = C$  for each subset  $C$  of the circle. The argument involves no conflation of models or labellings. It only assumes that the distribution is rotationally symmetric, and hence, that a rotationally symmetric continuous distribution is possible. In the dart throwing implementation, this amounts to assuming that it is possible to throw a dart, or construct a device to throw a dart, in

---

translation invariance, and consequently lack certain theoretical virtues. Parker 2012 gives the parallel argument against regular probabilities.

<sup>26</sup> Parker 2012 argues contrapositively from the assumption of regularity to the failure of rotation invariance, but again it is explicitly a failure of rotation invariance for a *single* probability function. Bernstein and Wattenberg 1969, Barrett 2010, and Pruss 2013 also discuss invariance for a single probability function. Of course, Benci et al. could claim that these are all careless glosses, but there is no need for such accusations if the arguments are taken at face value.

such a way as to yield a rotationally symmetric distribution. A dedicated regularist would have to deny that such a strictly symmetric distribution is possible. But that is a strong claim to make on the back of intuition, conceptual analysis, or theoretical virtues. It is at least conceptually possible that some perfectly symmetric set-up could produce a perfectly symmetric distribution. Benci et al. do not deny this; they only hint that the argument involves a conflation of two different models, and that is simply not the case.

#### 4. Next moves

If, as I have argued, Howson and Benci et al.'s replies fail to refute these three arguments, what more could they say in defense of regular and hyperreal probabilities?

Howson might respond by pointing to a merely instrumental role for hyperreal probabilities. He writes,

[T]he object there is not so much, or at all, to regard hyperreal probabilities as on the same footing as real-valued ones but to use the nonstandard universe simply as an aid to the standard theory by translating standard problems into nonstandard ones by means of the Transfer Principle, where they are often more tractable.... (2016)

Consequently, he might say, proponents of hyperreal probabilities will not be troubled by arguments from physical principles. However, this is not how philosophers typically use hyperreal probabilities. Hofweber (2014), defends hyperreal probabilities on conceptual grounds. His *Minimal Constraint* on probability measures is

**(MC)** If the chance of  $p$  is 0, then not  $p$ . If the chance of  $p$  is 1, then  $p$ .

"I can't help but to judge," he writes, "that (MC) is a conceptual truth about chance, given that 0 is the lowest and 1 the highest possible measure of chance. ... It is a conceptual truth about chance that an event which happens has a better chance of happening than an event which is conceptually incoherent." Lewis expresses a similar sentiment: "Zero chance is *no* chance, and nothing with zero chance ever happens" (1983, 176). Lewis and Hofweber do not use



infinitesimals to facilitate calculations, they just think that infinitesimals correctly represent the structure of chances in the real world (whether as a merely conceptual truth or a more realist metaphysical claim). Benci et al., on the other hand, champion infinitesimal probabilities in order to make better sense of what they consider to be conceptually possible scenarios, such as infinite lotteries. NAP models, they argue, have theoretical virtues over the de Finetti (1974) approach to infinite lotteries (which is essentially just to drop countable additivity) and even over the standard treatment of continuous sample spaces. To an extent, the *possibility* of calculation is one of their concerns, for it is one of the stated motivations for their generalized continuity axiom. But their primary motivation is not to simplify calculations. It is to find enlightening models, models that can give us a better theoretical handle on problematic hypothetical processes. So Howson's instrumentalist view of hyperreal probabilities is not in line with the philosophical literature.<sup>27</sup>

Still, Benci et al. seem inclined to a milder pragmatism. Some of their discussion suggests a general antirealism about probability models. “[T]here is no reason to assume,” they write, “that there is a unique best way to model certain infinite probabilistic situations...” Thus they might counter the arguments against regularity by claiming that, even if the space-time invariance of probabilities is sometimes mandated by plausible or useful principles, the best models over all might involve an infinitesimal deviation from such invariance. Or they might just argue that it is useful to apply various models to a given process, if only to better understand the space of possible models and their virtues and limitations.

Yet, as Benci et al. themselves point out, it could be argued that, “There is such a thing as physical chance. And it is a legitimate task of our mathematical models to track this property.” Plausibly, the chances for a given experiment have a definite structure. The outcomes in any sequence of die rolls or coin flips exhibit a distinctive and robust pattern, largely independent of the detailed circumstances or the observer's conceptions. It is one of the main goals of

---

<sup>27</sup> Bartha and Hitchcock (1999) is an explicitly instrumentalist exception.

probability theory to accurately characterize and explain such patterns. Benci et al. respond to such a realist viewpoint as follows:

But our models can only track physical chance in a mediated way. In order to describe a physical system and its behaviour, our probabilistic models have to select a sample space and label the point events (that is, establish a connection between reality and point events in the model). For finite sample spaces, the labelling does not matter; but for infinite sample spaces, different labellings can result in different probability assignments. All this induces a degree of relativity in probability values of events. (2016, 34)

Thus, according to Benci et al., any probability model with an infinite sample space will involve some arbitrariness, whether it is a standard Kolmogorovian model or a regular one. Their main concern in this passage is arbitrariness related to the choice of labelling and, for NAP models, the choice of an ultrafilter, but it suggests they might take a similarly noncommittal stance toward the choice between regular and space-time invariant models.

The problem with such a stance is that it appears incompatible with the goal of accurately modelling physical chances. For the kinds of experiments discussed here, a model cannot be both regular and space-time invariant. If our goal is to characterize the true structure of the chances in such experiments, we should take into account whether the chances are truly space-time invariant or regular (or neither). This leaves us little freedom to choose; either regular models are accurate or they are not, and the examples discussed here give us some reason to believe that, at least in those cases, they are not.

This brings us to another possible position, namely that of a moderate, pluralistic regularist who holds that, in cases where there is a strong argument from IP against regularity, the latter might fail, but otherwise it should hold. However, this position is awkward, especially for Benci et al. Their main application of NAP is to the de Finetti lottery with an infinite number of tickets, but their own urn argument suggests that such a lottery can bring regularity into conflict with IP or other plausible symmetry assumptions (provided there are cases where something like their renormalization step applies). To hold this pluralistic regularist position would mean holding that infinite lotteries are not regular when the specific conditions that

justify such a renormalization step hold, but they are generally regular otherwise. If we admit that regularity is false for certain selection mechanisms, why should we expect it to hold for others?

We can make this point more concrete. Suppose we have a lottery machine for which the renormalization step is valid, and suppose the moderate regularist admits that regularity fails for this lottery machine. Now let us add to this machine a component that detects which tickets are present in the urn. If one of the original tickets is removed, it applies a different selection mechanism for which no such renormalization formula applies. For this composite lottery machine, we cannot make Benci et al.'s urn argument. Will the moderate regularist then claim that regularity does hold for the composite machine? Surely, if the composite machine applies the same mechanism as the original machine when all the original tickets are present, then in that case it produces the same distribution as the original machine. Thus, such opportunistic regularism is generally untenable. Similarly, if we accept that there are realizations of the circle example where the distribution is fully rotation-symmetric and regularity fails, we should not expect that regularity holds whenever the distribution is not perfectly symmetric.

What this illustrates is that, if indeed we are concerned with accurately modelling the structure of objective chances, then the question of regularity turns not on theoretical virtues, but on the details of the probabilistic processes under study. If indeed there are cases where regularity does not hold, then (1) there is no sound and fully general argument for regularity, and (2) regularity is not needed to render such experiments conceptually coherent. At most, regular models boast certain theoretical virtues while lacking others, namely those of permitting invariance under various transformations. But if there are any facts about the structure of chances, the model should reflect those facts first, and desirable theoretical virtues only as accuracy permits. Of course, it may be difficult to determine what the most accurate model is in any particular case, but if we have good reason to believe that chances are not regular in certain cases, we can reasonably hypothesize that they are not regular in similar cases either.

## 5. Conclusion

We have reviewed three arguments that certain hypothetical experiments exhibit non-regular probabilities. If these arguments succeed, then regularity does not generally hold, and there is little reason to believe that it typically holds for other experiments, nor that we should demand it in our credences. Howson and Benci et al. have attempted to refute those arguments, but their refutations fail. Howson points out that Williamson's events are not in fact isomorphic, because one is a singleton while the other is a pair, but this misses the point. Howson is speaking of the abstract "events" of mathematical probability theory, which are sets, while Williamson is concerned with events in the ordinary sense of things that could happen. When Williamson says that his two coin flip sequences are isomorphic, he does not mean that they are subsets of a sample space that have a one-to-one correspondence, he means that they have all of the same physical qualitative properties, and this is true by hypothesis. Benci et al. claim that Williamson, Parker, and they themselves found their arguments on a conflation of different probability models. The symmetries between Williamson's coin flip sequences imply that they *can* be assigned the same probability in different models, but not that they must have the same probability in a single model, and likewise for Parker's point sets and their own lottery draws. But none of these arguments is in fact based on such a conflation of models. Williamson's and Benci et al.'s are both founded on the principle that qualitatively identical events in qualitatively identical circumstances should have the same probability, and Parker's is based on the plausibility of a perfectly symmetric continuous distribution. All claim, not that their parallel events can be given the same probability in *different* models, but that the parallel events will have the same probability in any *one* model, if that model is accurate. This is no mere slide.

The principle underlying the circle and urn arguments, that "isomorphic" events have the same probability, is not above dispute, but we have provided here a simple argument from more fundamental hypotheses. If (I) the laws of nature are space-time invariant, and (II) chances are determined by local qualitative circumstances and natural laws, it follows that qualitatively

identical events have the same chance, and should also be assigned the same credence insofar as rational credences track chance. One who insists on regularity must therefore deny either the space-time invariance of laws or the grounding of chance in laws and qualitative circumstances.

This leaves the regularist several options, including at least the following: One may take a more or less instrumentalist view that is more concerned with the theoretical virtues of regular probabilities than with accurately modelling chances. One may hold that regularity fails in the cases discussed but is still plausible in other cases, though we have seen that this is an uncomfortable position to hold. Or, one might simply deny IP, as well as the very possibility of a symmetric continuous distribution. Hofweber (2014), at least, prefers the latter move, and denies premise (II), that chances are determined by laws and local circumstances. But if regularity requires that so-called objective chance is in reality such a contextual matter, or that the laws of physics are not in fact space-time invariant, then the arguments for regularity should be regarded very sceptically.

## References

- Appiah, Anthony. 1985. *Assertion and conditionals*. New York: Cambridge University Press.
- Barrett, Martin. 2010. The possibility of infinitesimal chances. In *The place of probability in science*, eds. Ellery Eells and J.H. Fetzer, 65-79. Dordrecht: Springer.
- Bartha, Paul and Christopher Hitchcock. 1999. The shooting-room paradox and conditionalizing on measurably challenged sets. *Synthese* **118**: 403-437.
- Benci, Vieri. 1995. I numeri e gli insiemi etichettati. *Conferenze del seminario di matematica dell'Universita' di Bari* **261**: 1-29.
- Benci, Vieri, and Mauro Di Nasso. 2003. Numerosities of labeled sets: A new way of counting. *Advances in Mathematics* **173**: 50-67.
- Benci, Vieri, Emanuele Bottazzi, and Mauro Di Nasso. 2014. Elementary numerosity and measures. *Journal of Logic and Analysis* **6**: 1-14.

- Benci, Vieri, Leon Horsten, and Sylvia Wenmackers. 2013. Non-Archimedean Probability. *Milan Journal of Mathematics* **81**: 121–151.
- . 2016. Infinitesimal Probabilities. *British Journal for the Philosophy of Science*.  
<https://doi:10.1093/bjps/axw013>
- Bernstein, Allen R., and Frank Wattenberg. 1969. Non-standard measure theory. In *Applications of model theory to algebra, analysis, and probability*, Ed. W. A. J. Luxemburg, 171–85. New York: Holt, Rinehard and Winston.
- Boddy, Kimberly K., Sean Carroll, and Jason S. Pollack. 2016. De Sitter space without dynamical quantum fluctuations. *Foundations of Physics* **46**: 702–735.
- . 2017. Why Boltzmann brains do not fluctuate into existence in the De Sitter vacuum. In *The philosophy of cosmology*, eds. K. Chamcham et al., 288–240. Cambridge: Cambridge University Press.
- Carnap, Rudolf. 1950. *Logical foundations of probability*. Chicago: University of Chicago Press.
- . 1963. Replies and systematic expositions. In Schilpp (1963), 859–1013.
- De Finetti, Bruno. 1964. Foresight: Its logical laws, its subjective sources. In *Studies in subjective probability*, eds. H. Kyburg and H. Smokler, 93–158. Huntington, NY: Krieger.
- . 1974. *Theory of probability*, Vols 1 and 2. Trans. A. Machí and A. Smith. New York: Wiley.
- Di Nasso, Mauro, and Marco Forti. 2010. Numerosities of point sets over the real line. *Transactions of the American Mathematical Society* **362**: 5355–5371.
- Easwaran, Kenny. 2014. Regularity and hyperreal credences. *Philosophical Review* **123**: 1–41.
- Edwards, Ward, Harold Lindman, and Leonard J. Savage. 1963. Bayesian statistical inference for psychological research. *Psychological Review* **70**: 193–242.
- Hofweber, Thomas. 2014. Infinitesimal chances. *Philosopher's Imprint* **14**: 1–34.
- Howson, Colin. 2017. Regularity and infinitely tossed coins. *European Journal for Philosophy of Science* **7**: 97–102.
- Jackson, Frank. 1987. *Conditionals*. Oxford: Blackwell.

- Jeffrey, Richard. 1992. *Probability and the art of judgment*. Cambridge, UK: Cambridge University Press.
- Jeffreys, Harold. 1961. *Theory of probability*, 3rd Edition. Oxford: Clarendon Press.
- Kemeny, John G. 1955. Fair bets and inductive probabilities. *Journal of Symbolic Logic* **20**: 263–273.
- . 1963. Carnap's theory of probability and induction. In Schilpp (1963), 711–738.
- Lewis, David K. 1980. A subjectivist's guide to objective chance. In *Studies in Inductive Logic and Probability*, v. II, ed. R.C. Jeffrey, 263–293. Berkeley and Los Angeles: University of California Press.
- . 1983. Causation. In *Philosophical papers, Volume II*. Oxford: Oxford University Press.
- . 1994. Humean supervenience debugged. *Mind* **103**: 473–490.
- Parker, Matthew W. 2012. More trouble for regular probabilities. PhilSci Archive, <http://philsci-archive.pitt.edu/id/eprint/13063>
- . 2013. Set size and the part-whole principle. *Review of Symbolic Logic* **6**: 589–612.
- Pruss, Alexander. 2013. Null probability, dominance and rotation. *Analysis* **73**: 682–685.
- Schilpp, Paul Arthur, ed. 1963. *The philosophy of Rudolf Carnap*, The Library of Living Philosophers Vol. XI. Chicago: Open Court.
- Shimony, Abner. 1955. Coherence and the axioms of confirmation. *Journal Of Symbolic Logic* **20**: 1–28.
- Skyrms, Brian. 1980. *Causal necessity: A pragmatic investigation of the necessity of laws*. New Haven and London: Yale University Press.
- Stalnaker, Robert C. 1970. Probability and conditionals. *Philosophy of Science* **37**: 64–80.
- Wenmackers, Sylvia, and Leon Horsten. 2013. Fair infinite lotteries. *Synthese* **190**: 37–61.
- Wenmackers, Sylvia. 2011. *Philosophy of probability: Foundations, epistemology, and computation*. Dissertation, University of Groningen.
- Williamson, Timothy. 2007. How probable is an infinite sequence of heads? *Analysis* **67**: 173–180.

